TS PST THERMAL STABILITY (H/C RATIO)

$N = 905$

TS PST ACTIVITY (U/10^8 PLATELETS)

FIG. 1A

TS PST THERMAL STABILITY (H/C RATIO)

SULT1A1
☐ *2/*2
▨ *1/*2
■ *1/*1

TS PST ACTIVITY (U/10^8 PLATELETS)

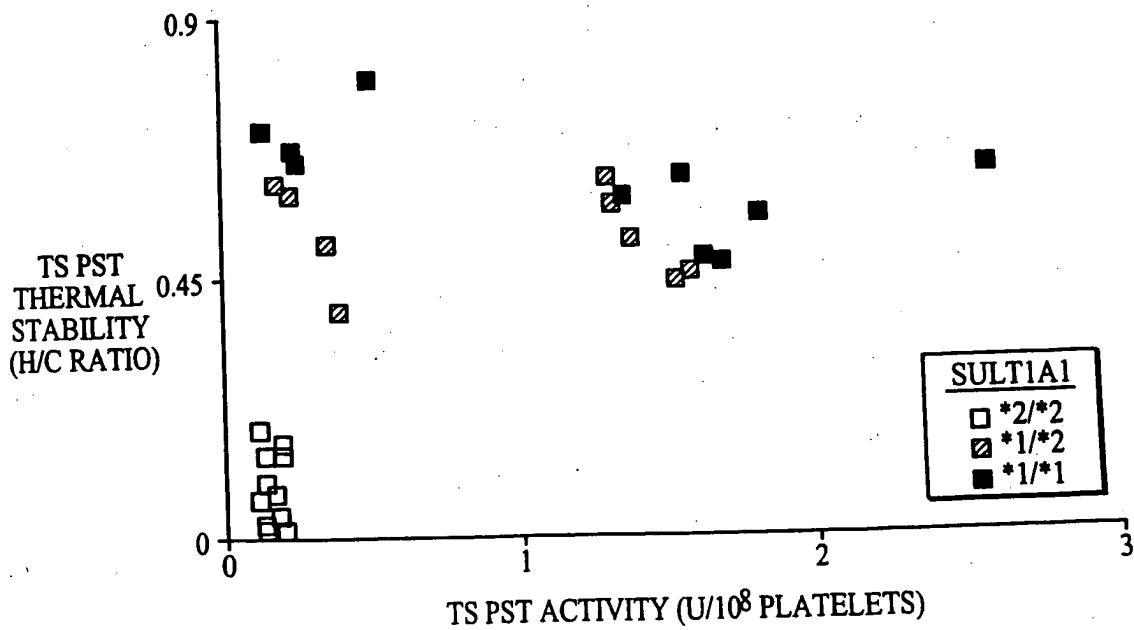## FIG. 1B

FIG. 2

FIG. 3

Appln No.: 10/668,800
Applicant(s): Richard M. Weinshilboum et al.
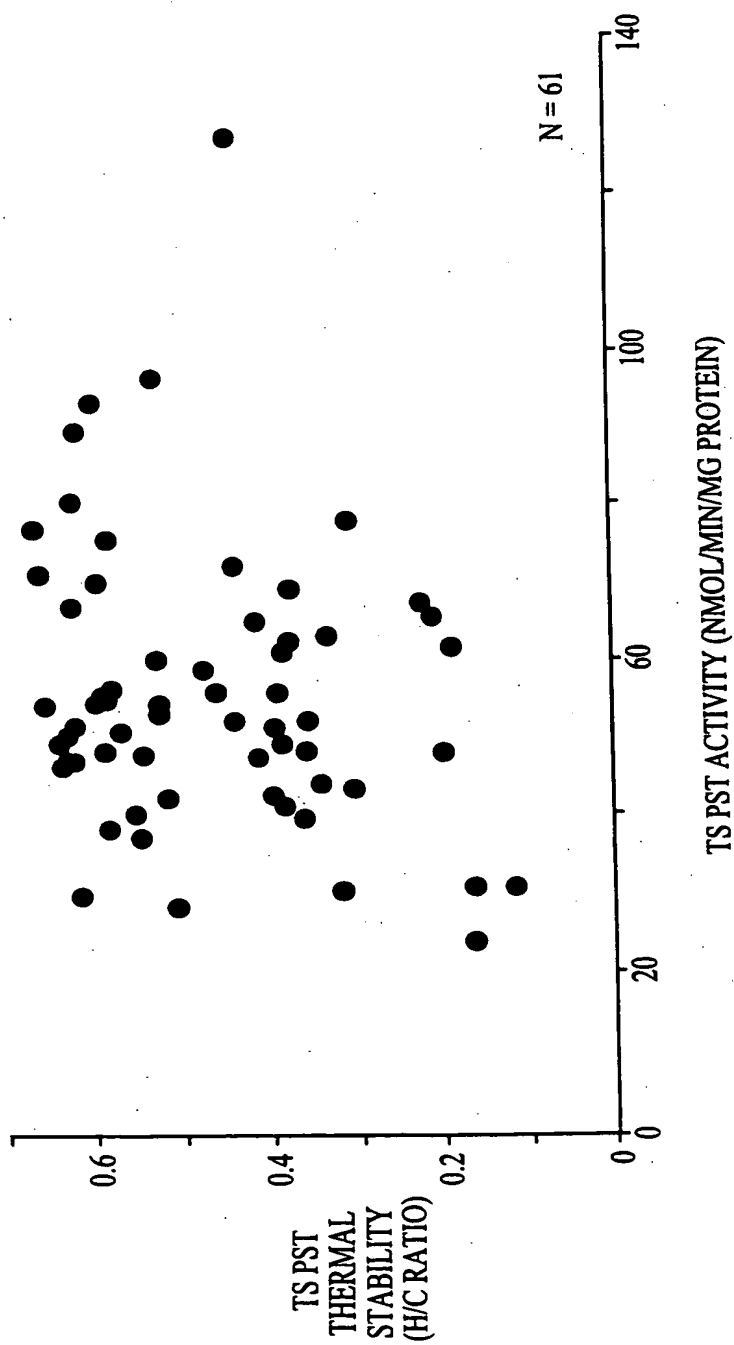SULFOTRANSFERASE SEQUENCE VARIANTS
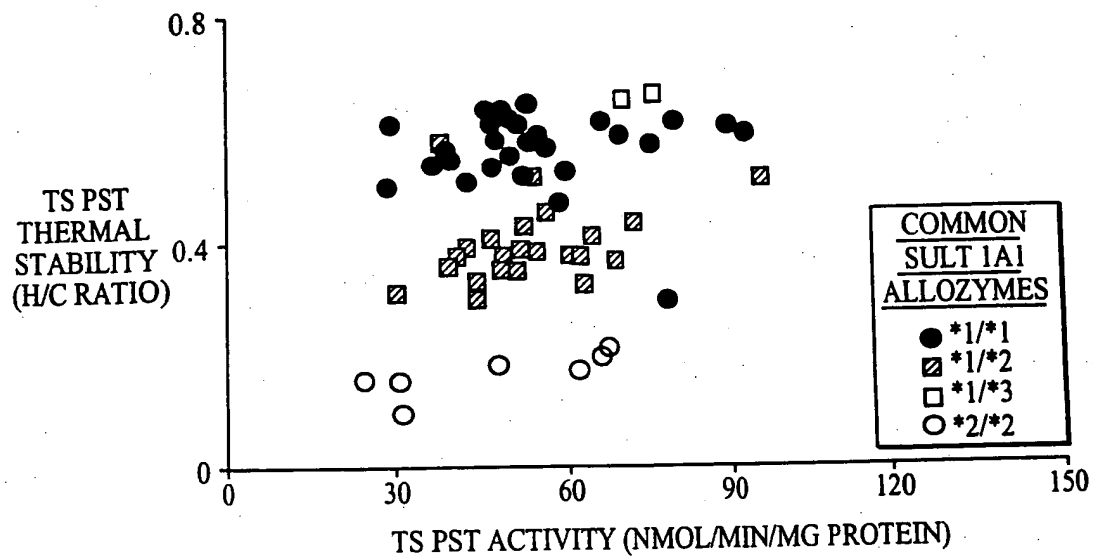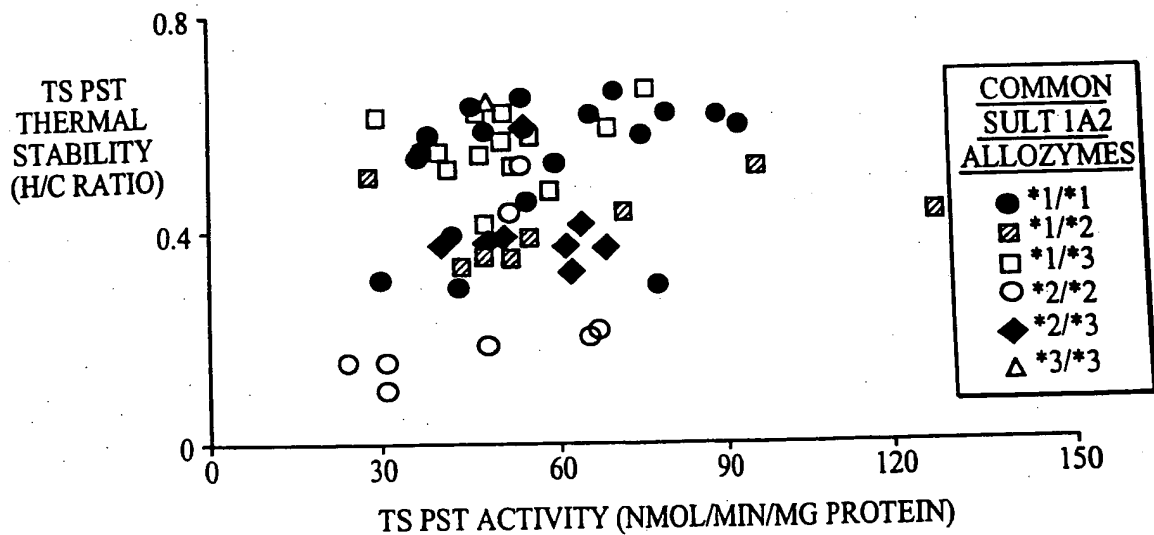
Page 4 of 13

FIG. 4A

FIG. 4B

```
  1 ttgctgccag ctgcctctcc otccttgtct cttacctgcc tgtgcctgg gacaggatga agcgggcc ttgtgttgcc ccaccctgg ctgttggcta
101 agagcccacg tgatctgcct gtgagaggag ttccttccgg aagaaccagg gcagcttotg occctagagg gcaatgcc tagctgagtg cagtccccg
201 gcccagcct gtccagctt tgggaagagg gtgccagtt gtgcaatcca gpccggggca gccgtgtcct gatcttggta ttcaggggtg agcctggagg
301 gggottgtga tgcctgacto tgtctccctc tctgcgccca tgccttggta gctgtgagc gtoactgctt tggggtgacct gatctggctg tgatggatga
401 gcacgggga aatagtgaa gactcggaat tagaagacgt gagtgggaat ctccctacco agtactgt gcctagggg aggtcatcaa aaacggtgg
501 cctcgtttct gcaggcacac tggatagcc tgtgtgagct ccgtgtcct aatccctcc agatactgt gttcactct tgttgccag gotggagtga
601 gacatcgacc tcagccgtt tccacgcttt tttttgtttt cagtgacoga gttcacctot tgttgccacg gattccagc gtgtgcaacc aggcttgact
701 tcttggctca ccgcaacctc cgcctcctgg gtcaagcga ttctcctgcc aagtagctgg tgtctgctc gtcatctgc tpctcggcc ccccaaagtg
801 aattttctat ttttagtaga gacaaggttt ctcoatgttg gtcaggtgg totcaaacto cgacttcag gtatotgcc taaggacaag tgaaaagtc atgaatgtag
901 ctgggattac aggagtgagc cacgtgcca ggcctctcc agctcttgg gtcctcaaga agaacaatt taaggacaag gttgtggggc tgcctttat
1001 gpagatttca tgcagagtaa agggactcac toaagaagag gaacgtgggg gtcctcaaga tcaagacgt gtgctcaa tttacaccca aatacaggtg
1101 gpgcttcttc aactaaagag aggagtatto gggtatattc cagggaaagt taaagattto agatcctag gtgaacota catcaaatac agggccatgt
1201 ttcctggag cgtcttggca otggtgggtg tacggttca ctcatcctat cttattggcc cttagcgcat ccccgtgtgg gtcgcacag tcttctcctc
1301 tgtgtctggt tggtcgtggc cagcttgtc ctcatcctat tttcaggga atctcacatt otcgcctctt gctagtagcc ccagtcacgg gtcogacagg
1401 atgtgaaact gctgcctggg attttgtatt cactgctac tcaagggaac otctgagggg aaggggtgga gotagtagcc catgcctaag ottccccac
1501 ttgttactag agtgcaatac aaagtcttag tcaagggaac gggtccatg gcctagaat atgagagcc totcctgtgc catgcctcag gtpacagct
1601 gpattcctca tgggacaga ggagggagga gggcatttct tggaagtott gggagttca cctttactda gatgttgtt taccttgtt gtgacagct dctotggtgt
1701 actcccacct gtccgottgc ototgaacto otgtgaacto agggaggga ccoooottcc agggccctg acctggggct gggttccaa acctgggct aggtcccag
1801 tgaacttgga cttaaagtg aggataaaga agaggagga cacctgcaac gggttgtgt tottotagga tdttctatca atgttctatca otgocctta
1901 ottgtctgtg gagcctctt oaaaccoagg caatgggact gcagtgtcct tgttctttct atccatgtc tgtctatgc oactgggca gaaacoctgg
2001 gggccatgaa gctggggctg actccccaggg actccaggg caatggggga agagocctgga gtgatgggc ctgggggggc ctgtggggp caotggtca gaacocctgg
2101 aatctgccca cacacatcc totagactgg cctgtggto togagggagc aggtcatocc agcttctcc tgcaabcct ccoctctctt gctgttca
2201 cacttcaag actggctggg agccagaggg tagttgcct gttcagggo gttcagggo agtoagggo oooagtgaoa gctcggaaa gtcgttggtt tggagtcaao
2301 attagagaga gcttgtctgt tgpctgootg tgcctgootg ccttgggcc ttccccttt cattcttotg ttttotctac agaaaotca ttocttcggc aagaaatot
2401 cagaaaaaaa aacctacaaa acctacaaa ccattgggca ocattgggco ccoaoacoc acooacoooa aaaaaaaaa gtcggaaa aagaaaaaa
2501 tttaagcttg totccagctg gcacactaag gagagtaatg gagggtaatg gagagtaatg
```

FIG. 5B

EXON IB

EXON IA

EXON II

EXON III

EXON IV

EXON V

EXON VI

EXON VII

EXON VIII

5501 ctcctttgcc aatccaagag atgagctggc ctgggccagg ctgtgtggtg ggttgagtct tctgccctg cagTGGTGCT ATGTTGCCCG
                                                                                        V Y V A R

5601 CAACGCAAAG GATGTGGCAG TTTCCTACTA CCACTTCTAC CACATGGCCA AGGTGCACCC TGAGCCTGGG ACCTGGGACA GCTTCCTGGA GAAGTTCATG
      N A K   D V A V   S Y Y   H F Y   H M A K   V K P   E P G   T W D S   F L E   K F M

5701 GTCGGAGAAG gtgggtttga tgggaggaag gaaagtgtgg agccgagggg tggtggctac aacctgtgt tggcaccct tgcctgcttc
      V E E K

5801 tccagTGTCC TACGGGATCCT GGTACCAGCA CGTGCAGGAG TGGTGGGAGC TGAGCCCGCAC CCACCCTGTT CTCTACCTCT TCTATGAAGA CATGAAGGAG
           S   Y G S W   Y Q H V Q E   W W E L   S R T   H P V   L Y L F   Y E D   M K E

5901 gtgagacac ctgtgaagt tcctacatg tgacaactgg gggggcgcac cctcacgggg tcaacaggc cctccttg gcagcccca

6001 cagaggccc ggattcccca tactgcctc ttgccccagc cctcccccgt aaggcccca acctgtagog ggtgtgcc accagaggcc aagaagacca

6101 gagtcagctg catggggcc cacggatcag agccaggacc ctattgctac accccccggga agggggcct gggggttgag

6201 tcaaggtcct cacccaggtag agccaagcc agtcatgat gggatcacag ggcagacago aattcttttct acccccgga tccgagccttc

6301 gggttggctc tgcaggtct ctaggcgagg gcctctgagg ttagagaagg gggaccccctt tactttcct gattcagcaa tccgagcctc

6401 caotgaggag ccctctgctg ctcacAACCC ATTCAAAAGA TCCTGGAGTT TGAGGGGCAC TCCCTGCCAG AGGAGACCGT GGACTTCATG
                         N P   K R E   I Q K I   L E F   V G H   S L P E   E T V   D F M

6501 GTTCAGCACA CGTCGTTCAA GGAGAATGAAG AAGAACCCTA TGACCAACTA CACCACCGTC CCCCAGGAGT TCATGGACCA CAGCATCTCC CCCTTCATGA
      V Q H   T S F K   E M K   K N P   M T N Y   T T V   P Q E F   M D H   S I S   P F M R

6601 GGAAAGgtgg gtgtggcca gtacgggggt ttgggggggt tggggcgagc agtgcagcago otcccctggg otcccctggg atgagactcc

6701 agccttgctc cctgccttcc cccccaggc ATGGGCTGGGG ACTGGAAGAC CACCTTCACC GTGGCGCAGA ATGAGCGCTT CGATGCGGAC TATGCGGAGA
                                    M A G   W K T   T F T   V A Q   N E R F   D A D   Y A E K

6801 AGATGGCCAGG CTGCAGCCTC AGCTTCCGCT CTGAGCTGTG AGAGGGCCTC CTGGGGAGGA TGGAGGCTGA GGTGGGAGGA TCATTTGAGC CCAGGAGTTC
       K G
       M A G   C S L   S F R S   E L

6901 AAGAATAAAA TATGAAATTGA GGGCCCTGGGA CGGTAGGTCA TGTCTGTAAT CCCAGCAATT CCCAGCAATT TGGAGGCTGA

7001 GAGACCAAACC TGGGCAAACAT AGTTGAGAATTC TCTTAAAAAA ATAAAAATAAA ATAAAAACCAA TTTTTaaaa gagaataaaa tatgattgtg ggccaggcat

7101 agtggctcat g 7152

FIG. 5C

-3729 ctctcctcc ttgtctctta cctgcctgct gctgggaca ggatgaagcg gggccttgt gttgcccaa cctgcctgt gttgcccaa cctgcctgt tggctaagag cccacgtgat

-3629 ctgcctgtga gaggagttcc ttcaggaaga acaggggcag cttctgccco gggggagccg tagagggcca atgcctagc tgagtgcagt cccccgccc cagctggtc

-3529 cagctttggg aagagggtgc coagttgtgc aatccaggcc ggggagccg tgtcctgatc ttggtattca gggctgagcc tggaggggc ttgtgatgcc

-3429 tgactctgtc totctctctg gcoccatgcc ttggtaggtg tgaggcgtca otgtttggg tgacctgato otgttgat gggctgagcc tggaggggc ggatgagcac gggggaaata

-3329 gtggaagact cggaattaga agacgtgagt ggccttggc ccoagcctcc ctaccocact cctgtcctg tgacctgato gacoaactt gttctgcag

-3229 gcacactgga tagcctgct ggagtccagt gtcctaato cctccagat actggtggcc actgtggcc tagggggt catcaaagac cagtgggaca tcgacctcag

-3129 cctgtttcca cgttctctgt tgtttttttt ttttgtgga gaagagttt oactcttgtt cactcttgtt oactcttgtt gctctgtctt gctcaccgc

-3029 aacctctgcc tcccgggttc aagcgattct cctgcctcag cctcccaagt agtcggatt acaggcgtgt gcaccagco gagtgaatg gctgtcctt ggctcaccgc ttctattttt

-2929 agtagagaca aggtttctcc atgttggtca ggctggtctc ttcaaggtga cttcaagtgta tctgcctgcc aaagtgctgg gattacagga

-2829 gtgagccacc gtgccaggcc ttctccagc tottgcaco tagcocgaa aacttaag sagtcatga agtagccag atttcctgca

-2729 gagtaaaggg actcactgaa gaagaggaac gtggggtcc toaagagt ttagcagaa acaattaag gacaagtgca ctttatgcc ttcttcact

-2629 aaagagggt atattcatga agagtgcagg aaaagtaaa gatttctcaa gctctcatgcc ctacaaggtg cacccaaata caggtgttcc tggagcgtc

-2529 ttgcactgg tgggtgtacg gtttcatatg ttactgattg tacagtgaga tootaggtga aacctacatc aataoagcg ocatgttgct totgttgt

-2429 cgoagccagc ttggtcctca toctatttt caggggactta ttgcccttg gcacatgoag otatttoag ttcccttctt otgtcatgt gaaactgctg

-2329 octgggattt totgttgtct tgctagcact otattaatct ototttctt oototttctt gtgcacocc otgctggtcc ggctggtttt cactcagtg

-2229 caatacaaag tctcaagtcaa gagggcctcc tgaagtttgc tgtggagata gtgggagbta gacotgccag tcatgggat tcctcaggg

-2129 cacagagag ggaggagggg cctgtggccc tagcaggga gcagcctcto ctctgcctgg aaatcccatg cccgcttgc ctctggctc

-2029 acgcaacct gggaaggctt gggagactca cctttactca gatgttgtt tacctgtctc gtgccaggt cctcctgtg ctttaaatag tgaggacaaa

-1929 gaacgaggag ggtgggggga tgcactcctt ccacggggc ctgtggctto caagcotcaa cctcctctgg totctgtgtg tggagcctcc ttcaaacca

-1829 tggaaagaaa agtacctgcc agggctgtg gttcttctag gttcttctag catgttctct cagtgtctg tgagtcccc aggggcct gagctgggg ctggctcooa

-1729 gggaaatggg actgcagtgt cctgttcttt tcttggttct atgatccat gtctgctcc accctgccc cttcactctg occacagca tcactccaga

-1629 ctgccttgt ggtoagagcc tgagtgcat gggctgctgg aggctgctgg gttgcactgg gcaggacoo coagcacctt caagactggc ctggagccag

-1529 caggtaggtg acctttccag ggctgccta ctctccaat cctcccccto tottgcctgg gtoaattaga gaaagcttgt ctttggagt

-1429 tcaggggcag gtcaggagcc cagtgacago cagtgacago accattggc accattggc accattggc ccttocccct ttcattctto tgtttotac oaacoctaco

-1329 acaccaaacc oagtcgtggc tttggagatc acttaagct tgtctccagc acttaagct tggcaaacta tagagaagct aggagggtaa ccccaccco oaacoctaco

FIG. 6A

FIG. 6B

EXON IB

EXON IA

EXON II

EXON III

EXON IV

Applicant(s): Richard M. Weinshilboum et al.
SULFOTRANSFERASE SEQUENCE VARIANTS

```
 572  tgggcacagt ggttcacacc cgcaatctca gtacttgggg aggctgaggt gggaagatcc cttgaagcca gaagttccag ataagtctct tccaaaaaaa
 672  aaacttagct gtgcatagtg gtgtgtgcct gtaataccag ttactccgga ggttgaggtg gaggatcat ctgagcctag gaggttaagg ttacagcgag
 772  ctatgatcac acoagtgoac tccaggotgg gtgcagaga acaactgtct caaaaaaoga tgaatagaaa gagtgtccca ccagtgcgt ggctcacacc
 872  tgtaattcca gcacttgaag aggctgagge agtggatca cctgagacta ggagtttgag atcagcctgg ccaacatgac aaacoccat ctctactaaa
 972  aatacaaaaa aattagccgg gcatggtgge aggcatctgt aatcccagct aatgggagg ctgaagcagg tgaaatgctt gaagctggga ggcagaggtt
1072  gtagtcagcc gagacctcac cattgcaccg cagcctggga caoaagagca aaactctgtc tcaaaaaaaa aagaaaaaa taaaaagcg gcagtggca
1172  ggggctggg cctgttgtgg ctcacgcctg taataccagc actttcggag gtcgaggtg gcagatcac caggttagg agtttgagat cagtctggc
1272  aaoatggaga aaccogtct ctactaaaaa tacaaaaatt agccaggcgt tggggcaggc gcagtaatc ccagtacto ggaggctga ggaggggaa
1372  tagcttgcac ctgggaggcg gtggttgcag tgagccgga ttgtgccact gtactccagc ctgggcaaca agagoagaca caacgagaca ttgtttcaaa caaacaaat
1472  aaatatttta aaaggtttgc cacctgggtg gtcaacogct gtaatgccag catttggga ggccaagatg ggtggacgo ttgagctcag gagttcoga
1572  ccagccagg aaacatgggg agactccato tctataaaag tcagcaggc atgtggcat agcgctatag tccaagctac tcaaagtct
1672  aaggttggag gattgcttga gcctgggagg tcaaggtgo agtgagccat tgatatt actatagaat atgatgtct
1772  ggtctttttt ctcttttctc tttttttgaa gccctagagt occcccccc cttggtcct gttcttctgg gagtgcgcc tcatcttcca
1872  ggoatgtgaa tcaggggaag gaggccccgg gaggccccg otcagccag atgagcgaaa gtagagggtg ggtcctcaca gaccacaaa
1972  cgggggggac tgtctgcctt ttgctctgat gaccaaaaac tcggggtaga cctaagaaag gtagagggtg agattatatt actatagaat atgatgtct
2072  atttggtggt gtgggaaca tgcctggtgg agcatgcctt gtccagato atcctgactt ggccctaact gtgccgtgag gtgccgtgag tctcaaaca agtcagcct taagcctcag
2172  cagggaccag gcaggactt ggcttttgag tgoctggtg caggttcag atcctgactt ggccctaact gtgttotgcc cotgcagGTG GTCTATGTTG  ◄— EXON V
2272  cttcctcctt tgcaaaocca agagatgago aggcctgggg caggtctgt caggttttga otgggttga gtcttotgcc cotgcagGTG GTCTATGTTG
                                                                                                    V  V  Y  A
2372  CCCGGCAACGC AAAAGGATGTG GCGGTTTCCT ACTACCACTT CTACCACATG GCCAAAGTGT ACCCTCACCC TGGGACCTGG GAAAAGTTCC TGGAGAAGTT
      R  N  A   K  D  V    A  V  S  Y   Y  H  F    Y  H  M    A  K  V  Y  P  H  P   G  T  W     E  S  F  L    E  K  F
2472  CATGGCTGGA GAAGgtgggc ttgatgggg gaggaagt gtgggtaa gggtggtgg otacaacgca cagcaacct gtgtcggcac ccctgcccg
      M   A     E  V
2572  cttctccagT GTCCTATGGG TCCTGGTACC AGCACGTGCA AGAGTGGTGG GAGCTGAGCC GCACCCACCC TGTTCTCTAC CTCTTCTATG AAGACATGAA
           S   Y  G      S  W  Y      Q  H  V    Q  E  W  W    E  L  S    R  T  H    P  V  L    Y  L  F  Y   E  D  M   K
2672  GGAAGgtgaga ccgccttga tgcttccctc cacgtgacac ctgggggcag gcacttcaca gggcctcaca agggcaccc cctgggcggc
      E
```

EXON VI

FIG. 6C

```
   1 acctctgcct cctggttcca agcaatcctc cttcctcacc ctccagagta gctgggatta
  61 cacgcgcctg ccaccgcgcc tggcctaatt tttgtatttt tagtagagat gggggtttcc
 121 aaccatgttg gccaggctgg tctccaaact cctgacctca ggtgatcctg cccacctaag
 181 cctcccaaaa tgctggtatt acaggcatga gccaccgtgc ccggcctaaa taattaataa
 241 aataatggac gatgggtgcc ttctactgag ctcccggtaa ttgtgagtga gtagaggact
 301 tgccctgggg acattcagtg acctgctggg tgttgctgag ctgtgaggaa gttcaggtct
 361 ggctgcagtg gtgaggctgt gactcaatca atcactgctg atgctcccag gacctgcacc
 421 agcttagtcc tagggcaag gattttaact gtccacctca gtctcttcat ttgtaagatg
 481 caaataacag tcacccctgc ctcatgggat ggagctgtgt aatgcccgca acagtgcctg
 541 ctgcatagag gggttgctgc cagctgcctc tccctccttg tctcttacct gcctgctgcc
 601 tgggtcagga tgaagagcag ccctgtgtt gcccccaccc tggctgcctg ctaagggccc
 661 atgtgatctg cctggcagag gagtttcttc aggaagaacc agggcagctt ctgcccctag
 721 agggccaatg cccttggtga gtgcagtccc ctggccccag cctggtccac ctctgggaag
 781 agggtgccca gttgtgcaat ccaggcccag gcagctgagc cctcatctca gcatgcaggg
 841 cggatactgg agggggcttg tggcatctga ctctgtatct cctacctgcc cctctccttg
 901 gtacctgtga gaagtcactg ctttggggag acctgatctg gctgtgccag atggacactg
 961 agaaagaagt agaagactca gaattagaag aggtgagtgg gctttggtgg cgggctccct
1021 accccactcc ctgccctggg ctgcctgtga ccacactgct tgcctctgca ggcacactgg
1081 acagacctgc tggagacctg atcctcagtg tccttacccc ctcctacctc ttttctgtgc
1141 cacctgctgt gggtccagca ggttttact tgagtacaat aaaaagtctg agtcaagggt
1201 gccttatggt ggatgctgag gggaggggcg gagctagtag cccaaggtcc tgccagtcac
1261 ggggcttcct caggggcaca gaggaggcag gaggggcccc tggccctagc acgtgaacag
1321 cttctactct gcctggaaac cccatgcctc agctttcccc tacttgcctc tgagctcatg
1381 caattcttgg aagcctggga gacttacctt gaaattgaat gcaaatagga caaagaccaa
1441 ggaggatggg gggatgccct ccttccacgg ggccctgtgg cttccaagtc ttaatctcct
1501 ctagtctctt gtctacggag cctccttcaa acccagggaa agaaaagcac ctgccagggt
1561 tgttttttctt ctaggatctt ctattgatgc tctgtgaggt cccccaggag ccatgaagct
1621 agggctggct cctagggcaa tgggactaca gtgtccttgt cctttcttat tctttctgtt
1681 cttcttttct ttcttttttt tttttttttt tttttttgag acagagtctc actctgttgc
1741 ccaggctgga gtgcagtggt gtgatcttgg ctcactgaaa cctccgcctc ctgggttcaa
1801 gtgattctct tgcctcagcc tcctgagtag ctaggattac aggtgcccgc catcatgccc
1861 agctaatttt tgtattttta gtagagacag ggtttcacca tgttggccag cttggtctcg
1921 aactcctgac ctcaggtgat cctgctgcat cgacctccca aagtactggg attacaggcg
1981 tgagccacca cgctcagcct ctttcttgtt ctatatgtcc atgctctgct ccacttctgc
2041 cccttcactc tgccccacac atcactccag actggccttg tggtcagagc ctggaatgcc
2101 tgggctgctg ggggcctgtg gactgcactg ggccagaacc cctgccgcct tcaagactgg
2161 cctgtagcca gcaggtaggt gactttttccc aggccggcct atcccacctt tcccctccac
2221 tcactcacct cccttgcctg ggtcaattag agaaagcttg tcggccaggc atggtggctc
2281 atgcctgtaa tctcagcact ttgggaggcc gaggcgggcg gatcatctga gctcaggagt
2341 ttgagaccag cctggccaac atggcaaaac cccgtctcta ctaaaaatac aaaaattaac
2401 cggatgtggt ggtgtgcacc tgtaatccca gctactcggg aggctgaggc agaagaatcg
2461 cttgaaccca ggaggggggag gttacagtga gcggagatcg tgctactgca ttgcagcctg
2521 ggcgagagag cgagtctcca tctcacataa aaaaaagaaa aagaagaaa gcaagcttgt
2581 ctgttggcct gccctgcagg gtggagttca gagggaaggt caggagccta gtgacagctc
2641 aaaaaaaaaa aaacccaaat accaatgttg gccccttttg cctttcattc atgtgttttc
2701 tatacactaa actcacatat tgggtttgca gatcactcca agcttggctg gagctgtggt
2761 ggtaaggagg gtaatagaga agcttcccca ccctcaaccc caccccttcc ttcctggagt
2821 tcccagccct gactttagat ccctcccaca ctggaccttc aaaaccctca gggcagagag
2881 cagccctaca ctccctacac cacacccata ctcagcccct gcaggcaagg agagaacagg
2941 tcaggttccc gagagctcag gtgagtgaca cgttggaatg gcccagggca ccttcaccct
3001 gctcagcttg tggctccaac attctagaag ccgaggcctc tgccatccct gccctttccc
3061 atggatattc catttcaatt agacaaccca gcctggccgg aacccccctg cgttccttct
3121 tttcctttgt gtatttttga gacagggtgt tgctccgtca cccaggctgg agtgtagtgg
3181 gatcctggcc cactgcagcc tcaaattcct aggctgaggc aatcctgccg cctcagcctc
```

FIG. 7A

```
3241 ctgagtagct ggggttacaa gagcaagcca ccacacccag ctaattttga aaaatatttt
3301 ttgtagagga gaggtcttgc tttgttgtcc aggttggtct caaactccag ggctcaaggg
3361 atcctttccc gttggcctcc caaggctctg ggattacagg cgggagtcac cctgcctggg
3421 cccctccttt tgatgagtca tcagttttca ttcccgcacg aggctctagc ccctggtacc
3481 agcttagttg ctcaatgggc tgtgtttgtt ctggagccca gatggactgt ggccaggcaa
3541 gtggatcaca gacctggccg gcctgggagg tttccacatg tgaggggcat gagggggggct
3601 caaggagggg agcatcgcgg agaggagcgc actgggtgga ggctgggggt cccagcagga
3661 aatggtgaga caaagggcgc tggctggcag ggagacagca caggcaggcc ctagagcttc
3721 ctcagcacag ctggactctc ctggagacct tcacacaccc tgatatctgg gccccgcgct
3781 acgagggtgc tttcactggt ctgcactatg ccccaggccc tgggattttg aacagctctg
3841 caggtgactg aaaggtgcgg ccaggctggg gaacgacctg gtttcagccc cagccccgcc
3901 actgactgac tttgtgagtg cgggcaagtc actcagcctc cctaggcctc agtgacttcc
3961 ctgaaagcaa aaactctgca aaggggcagc tgggtgctgg ctcacacctg taatcccagc
4021 actttgggag gctgaggtag acaaatcact tgaggccagg agtctagac cagcctggcc
4081 aacatggtga aaccccatct ctactaaaga aaaaaaaaa ttagctgagc atggttgtac
4141 atgcttgtaa tcccagctac ttgggatgcc gaggcgggag gattgcttga acccaagagg
4201 tggagtttgc agtgagctga gattgtgcca cactgcactc cagcttgggt gagagtgaga
4261 ctccatctca aaaaaaaaaa aaaaagaga gaatcccact ttcttgctgt tgtgatggtg
4321 gtaagggaac gggcctggct ctggcccctg atgcacgaac atggagctga tccaggacac
4381 ctcccgcccg ccactggagt acgtgaaggg ggtcccgctc atcaagtact ttgcagaggc
4441 actggggccc ctgcagagct tccaagcccg acctgatgac ctgctcatca acacctaccc
4501 caagtctggt aagtgaggag ggccaccgac cctctcccag gcggcagtcc ccaccttggt
4561 cagcaaggtc gtgccctcag cctgctcacc tcctatctcc ctccctctcc aggcaccacc
4621 tgggtgagcc agatactgga catgatcctc cagggcggcg acctagagaa gtgtaaccgg
4681 gctcccatct acgtacgggt gccccttcctt gaggtcaatg atccagggga accctcaggt
4741 gcatggctgg gtcctggggg taagggaagt ggaggaagac agggctgggg cttcagctca
4801 ccagaccttc cctgacccac tactcaggc tggagactct gaaagacaca ccgcccccac
4861 ggctcatcaa gtcacacctg cccctggctc tgctccctca gactctgttg gatcagaacg
4921 tcaacgtgag gccggctca atggttcaca cctgtcatcc cagtttgaga ctgaggaggg
4981 aggatccctt gaaggcgaga gatggagacc agcctgggca acattgctgt agagatgaca
5041 tcccatctct acaaaaataa aattaacaac ctggtatggt ggcatagact gttcccagtt
5101 acttaggagg ctcagcgggg aggactgttt atgcaaatag gaagctgcaa tgagccctga
5161 tgatcctgct gctgcactcc agcctgggca acacagcaaa accatctcta cgaaaaaaaa
5221 agttcccact gactggcaag gaaagccagg aagggggggct caggtgccct ctcagccatg
5281 tacctgttct tctggaaggg cctcctcgct tctgccaggc tcatcacatc ttttttttttt
5341 ttgagacaga gtcttgctct gtcaccctgg ctggagtgca gtggcatgat ctcagctcac
5401 tgcaacctcc gcctccccag ttcaagtgat tctcctgcct cagcctcctg agtagctggg
5461 attacaggcg tgtgctacca cacccggcta attttgtat tcttttttagt agagacgggg
5521 tttcaccatg ttggtcaagt ggatctcaaa ctcttgacct tgtgatcctc ctgcctcgac
5581 ctcacaaagt gctggaatta caggcgtgag ccaccgcgcc tggcccttttt tttttttgag
5641 acagtttcac tcttgttgcc gaggctagag cgcaatcgtg tgatctcggt tcactgcaac
5701 caccgcctcc tgggttcaag caattctcct gcttcagcct cccaaggagc tgggattaca
5761 ggtacctgcc accacgcccg gctaattttg tattttttagt agagatgggg tttcaccatg
5821 ttggtcaggc tggtcttgaa ctcctgacct caggtgatct ggccaccttgg cctcccaaag
5881 tgccgggatt agaggcatga gccaccacgc ccagccttca tcacatcttg agagaggaca
5941 ctgtctgcct cttgctctga tgagggtctg atgcaaagga tagtgagtct ctacagtgca
6001 cacttaagaa aggcagcatg tgggtgctca caggtcaggc ggaggagggg gagctggtgg
6061 ggaccaggca tgccttgctc cagatcagga tatgatggca ttggtgcaga ttatattagt
6121 atagaatatg gtctcaggaa ccaggcagga ctttggcttc cgagcagggt tcagatccca
6181 gcttggccct acctgtgcag tgagatctca agcaagtcag cctctaagcc tcaggttcct
6241 ccttttgccag ttcaacagat gagctggcct ggggtgggct gtgtggtgat ggtgctgggg
6301 ctgggtcctc tgcccctgca ggtggtctat gttgcccgaa acccaaagga cgtggcggtc
6361 tcctactacc attttccaccg tatggaaaag cgcaccctg agcctgggac ctgggacagc
6421 ttcctggaaa agttcatggc tggagaaggt gggcttgact ggaggaagga gggtgtgaag
6481 ccgagggggtg gtggctataa cgtacagcaa ccctgtgtcg gtgcccctg cccgcttctc
6541 tactgtccta cgggtcctgg taccagcacg tgcaggagtg gtgggagctg agccgcaccc
6601 accctgttct ctacctcttc tatgaagaca tgaaggaggt gagaccgact gtgatgcttc
```

FIG. 7B

```
6661  cccccatgtg  acacctgggg  gcaggcacct  cacagggacc  caccaaggcc  acccagcccc
6721  gtccctgggc  ggctcccaca  gcaagcccgg  attccccatc  ctacctccct  ggcccaggcc
6781  cccccactgc  agccccacct  ggcagcaggc  tcggcacagc  tttcatcttc  tgcacctgag
6841  tcagctgcat  gggtggccac  ggatcagata  cttagtccta  ttgcttatcc  tcaccaaagg
6901  gtgtgccacc  cagggccaca  gtcatggaag  aagaccatcc  cggtcctcac  ccatagcgc
6961  caagccctgt  tcatgatggg  atcacagggc  agagatcaat  tcattttact  ccajtgacta
7021  gggccccagg  ggttgaggct  ctttgggggtt  tctaggggaa  gtggccagat  cccctctgag
7081  gttagagagg  gggacccgtt  ttgtttttgct  ccactgagga  gccctctgct  gctcagaacc
7141  ccaaaaggga  gattcaaaag  atcctggagt  ttgtggggcg  ctccctgcca  gaggagacca
7201  tggacttcat  ggttcagcac  acgtcgttca  aggagatgaa  gaagaaccct  atgaccaact
7261  acaccaccgt  cccccaggag  ctcatggacc  acagcatctc  ccccttcatg  aggaaagtg
7321  ggtgctggcc  agcacggggg  tttggggcgg  gtgggagcag  cagctgcagc  ctccccatag
7381  gcacttgggg  cctcccctgg  gatgagactc  cagctttgct  ccctgccttc  ctcccccacc
7441  catggctggg  gactggaaga  ccaccttcac  cgtggcgcag  aatgagcgct  tcgatgcgga
7501  ctatgcggag  aagatggcag  gctgcagcct  cagcttccgc  tctgagctgt  gagaggggct
7561  cctggagtca  ctgcagaggg  agtgtgcgaa  tctaccctga  ccaatgggct  caagaataaa
7621  gtatgatttt  tgagtcaggc  acagtggctc  atgtctgcaa  tcccagcgat  ttgggaggtt
7681  gagctggtag  gatcacaata  ggccacgaat  ttgagaccag  cctggtaaaa  tagtgagacc
7741  tcatctctac  aaagatgtaa  aaaaattagc  cacatgtgct  ggcacttacc  tgtagtccca
7801  gctacttggg  aagcagaggc  tggaggatca  tttcagccca  ggaggttgtg  gatacagtga
7861  gttatgacat  gcccattcac  tacagcctgg  atgacaagca  agaccctccc  tccaagaaa
7921  ataaagctca  attaaaataa  aatatgattt  gtgttcatgt  agagcctgta  ttggaaagga
7981  agagaaactc  tgagctgaaa  gagtgaatgc  ccggtggggc  cacatatggt  cacctctccc
8041  ccagccttca  gctccccagg  tcaccatatc  tggggagggg  agaagggttt  ggagaagtaa
8101  aacccaggag  atgtgtggag  gggggatgtc  tgtttaatcc  cagcacatcc  tctgctgtcc
8161  tgccccaaga  tggtggagga  cgtcgagtcc  gccgggcagc  gtcacttttt  cttgggctcc
8221  ttagaagcta  ccaggtacct  ctgggccaca  ctgagatgag  gggagtagcc  gcctgcatag
8281  gaggtgtctt  caaacaggat  agtatagtcc  ctcctggggg  ttgtgggggt  aggtggccaa
8341  ggaagggtag  aggagcaagc  ccccgggggct  ggttgtcaac  tcactttgtt  ggctggaatt
8401  ggttgtaact  tgaccacctc  gggcaggatc  ccactgctca  tccccaa
```

FIG. 7C